基于迁移学习的多尺度特征融合牦牛脸部识别算法

陈占琦[1]，张玉安[1*]，王文志[1]，李 丹[1]，何 杰[1]，宋仁德[2]

（1. 青海大学 计算机技术与应用系，青海西宁 810016；2. 青海省玉树州动物疫病预防控制中心，青海玉树 815000）

摘 要：牦牛个体身份标识是实现个体建档、行为监测、精准饲喂、疫病防控及食品溯源的前提。针对智慧畜牧智能化、信息化等养殖平台中动物个体识别技术应用需求，本研究提出一种基于迁移学习的多尺度特征融合牦牛脸部识别算法（Transfer Learning-Multiscale Feature Fusion-VGG， T-M-VGG）。以预训练的视觉几何组网络（Visual Geometry Group Network，VGG）为骨干网络构建基于迁移学习的卷积神经网络模型，获取其Block3、Block4、Block5输出的特征图，分别用F3、F4、F5表示，将F3和F5经过三个不同膨胀系数的空洞卷积组成的并行空洞卷积模块增大感受野后，送入改进的特征金字塔进行多尺度特征融合；最后利用全局平均池化代替全连接层分类输出。试验结果表明，本研究提出的T-M-VGG算法在194头牦牛的38,800张数据集中识别准确率达到96.01%，模型大小为70.75 MB。随机选取12张不同类别牦牛图像进行面部遮挡测试，识别准确率为83.33%。本算法可以为牦牛脸部识别研究提供参考。

关键词：牦牛；脸部识别；迁移学习；特征金字塔；T-M-VGG

中图分类号：TP391;S823.8+5　　**文献标志码：**A　　**文章编号：**SA202201001

引用格式：陈占琦，张玉安，王文志，李丹，何杰，宋仁德. 基于迁移学习的多尺度特征融合牦牛脸部识别算法[J]. 智慧农业(中英文), 2022, 4(2): 77-85.

CHEN Zhanqi, ZHANG Yu'an, WANG Wenzhi, LI Dan, HE Jie, SONG Rende. Multiscale feature fusion yak face recognition algorithm based on transfer learning[J]. Smart Agriculture, 2022, 4(2): 77-85. (in Chinese with English abstract)

1 引 言

牦牛素有“高原之舟”之称，能够带动青藏高原地区牧民脱贫致富。但目前藏区牦牛养殖仍处于传统模式，缺乏信息化、智能化管理设备的广泛应用。对于中小规模养殖户来说，牲畜识别的主要方法是打耳标、颜色标记、热铁烙印等传统方法。耳标通过穿孔的方式安装在动物的耳朵上，会有细菌感染的风险，相互摩擦也会导致标签丢失。传统接触式识别技术不能为识别过程提供持久的保障。

目前国内外研究人员利用卷积神经网络（Convolutional Neural Networks， CNN）在牛、羊、猪等动物个体识别和行为识别方面开展了大量研究，但是对牦牛识别研究较少。秦兴和宋各方[1]基于双线性CNN，利用视觉几何组网络（Visual Geometry Group Network，VGG）作为特征提取器对200头猪的2110张照片测试，面部识别准确率达到95.7%，试验环境下运行时间为4 d，模型体量较大，在实时性应用场景中存在一定局限性。何屿彤等[2]基于改进的YOLOv3（You

收稿日期：2022-01-02

基金项目：青海省科技计划项目（2020-QY-218）；国家现代农业产业技术体系资助（CARS-37）

作者简介：陈占琦（1996—），男，硕士研究生，研究方向为智慧畜牧。E-mail：1046788801@qq.com

*通信作者：张玉安（1981—），男，博士，教授，研究方向为进化计算、智慧畜牧。E-mail：2011990029@qhu.edu.cn

Only Look Once）进行猪脸检测识别，虽然提高了模型精度，但是仍存在小样本检测边界定位问题。刘忠超和何东健[3]参考手写体字符识别CNN（LeNet-5）对奶牛发情行为识别，准确率为98.25%，漏检率为5.8%，能够实现奶牛发情爬跨的实时监测。杨秋妹等[4]使用GoogleNet对猪只头、背、尾部分类识别进而判断饮水行为是否发生，识别准确率为92.11%，有效地提高了养殖管理效率。张宏鸣等[5]提出了一种融合空间信息的注意力机制的羊脸识别模型，在开集验证中识别率为88.06%。魏征[6]采用基于局部保持投影的2D线性判别分析算法，研究全局和局部特征相结合的不完美牛眼虹膜图像识别问题，但对牦牛来说虹膜采集实施过程存在不便性。何东健团队[7,8]利用CNN提取奶牛背部和躯干特征，但适用于身体特征明显的牛只识别。陈争涛等[9]采用基于迁移学习的并行CNN牦牛脸识别算法，识别准确率达到91.2%，训练时间为2 d，但并行的基于迁移学习VGG16网络会增加模型存储成本。Hansen等[10]和Marsot等[11]使用CNN对猪只进行面部识别，识别准确率分别为96.7%和83.0%。Kumar等[12]研究基于深度学习的牛鼻纹识别，准确率为98.9%。Jung等[13]利用CNN对牛进行声音分类和行为分析，准确率为94.1%。Salama等[14]采用贝叶斯（Bayesian）优化寻找最佳CNN进行羊脸识别，准确率为98%。鉴于人脸识别和动物识别任务的相似性，已有研究将人脸检测与识别技术迁移到动物检测和识别的相关工作中。Wada等[15]利用特征脸（Eigenfaces）算法对10头猪进行识别，准确率为77%。Rashid等[16]以迁移学习为基础，训练一个能够找出人脸和动物面部特征相似性映射空间的网络来实现检测，由于牦牛面部毛发影响，给检测和识别过程带来挑战。虽然虹膜、鼻纹以及声音识别具有唯一性、稳定性的特点，但设备安装成本过高，且牦牛在采集虹膜和鼻纹过程中不易被控制，因此不适用于高原牧区的中小规模养殖场。牦牛体毛颜色特征差别很小，存在季节性退毛，因此很难对身体躯干部位展开识别研究。虽然以上研究中提出的非接触生物特征识别技术实施起来存在一些困难，但是利用捕捉到的面部图像进行识别具有一定优势。本研究参考以上文献中的方法，结合迁移学习训练速度快、VGG16提取特征能力强的优势，使用迁移学习结合特征金字塔来实现牦牛个体面部图像识别。

2 材料与方法

2.1 数据集采集

数据采集地点为青海省玉树藏族自治州，数据集中包含194头牦牛脸部信息。在试验过程中将每头牦牛视为一个类别。用连续两天的上午完成数据采集，对每一头牦牛拍摄大约2 min时长的视频，拍摄设备为GoPro8，画面分辨率为1920×1080，并将每一段视频转换成图像帧。部分数据集如图1所示。

2.2 数据集处理

为避免视频中连续帧之间相似性过高，使用结构相似性（Structural Similarity，SSIM）算法[17]进行处理。SSIM算法通过计算两幅图像之间的方差、协方差和平均强度来判断两幅图像是否相似，从而筛选淘汰掉相似度较高的图像。

为平衡各类别样本间的分布，增强模型的鲁棒性，对数据集中样本数量较少的类别采用数据增强扩充策略。

首先，将处理过的数据利用Python图像处理工具裁剪出分辨率为512×512、320×320的图片，将不同大小的图片缩放成同一分辨率256×256。

其次，利用OpenCV的相关函数随机增强图像数据：顺时针旋转角度为5°和10°，亮度降低参数设置为0.85、亮度增强参数设置为1.3，椒盐噪声和高斯噪声系数分别设置为0.15和0.2。

最后，对于样本数量明显过多的类别，随机剔除一部分样本，验证集的每一个类别数量保持相同。经过整理，数据集共计38,800张，对每个类别的样本顺序编号，其中训练样本31,040张，



图 1　牦牛脸部识别部分数据集

Fig. 1　Part of data set of the yak facial recognition

占比 80%，每个类别训练数量保持在 170～210 张；验证样本 7760 张，占比 20%，每个类别的验证数量保持在 40 张。其中训练集和验证集样本 ID 互斥。

2.3 试验条件

试验配置环境如下：Ubuntu 18.04.5 LTS 64 位系统，GeForce GTX 1080Ti 图形处理器（Graphics Processing Unit，GPU）加速，CUDA 和 CUDNN 版本分别为 10.0 和 7.4.1.5，编程语言为 Python3，基于 Tensorflow 的 Keras 深度学习框架完成试验训练及验证。

2.4 T-M-VGG 网络构建

基于 CNN 的迁移学习能够把数据量充足的源域所学到知识迁移到数据体量较小的目标领域，提升目标领域任务的学习效果[18,19]。本研究以 VGG16 网络的迁移学习为基本架构，在此基础上增加并行的空洞卷积模块和改进的特征金字塔结构完成试验。

2.4.1　并行的空洞卷积

空洞卷积可以在不改变参数量大小的同时增加感受野。空洞卷积计算如公式（1）所示。

$$Out(x,y)=\sum_{m}^{h}\sum_{n}^{w}Input\begin{pmatrix}x+dr\times m,\\ y+dr\times n\end{pmatrix}\times K(m,n) \qquad (1)$$

其中，$Input(x,y)$ 为输入特征层；dr 为空洞率；$K(m,n)$ 为核函数；h 和 w 为特征图的大小。

假设空洞卷积的卷积核大小为 k，空洞率为 dr，则等效卷积核大小 k' 如公式（2）所示。

$$k'=k+(k-1)\times(dr-1) \qquad (2)$$

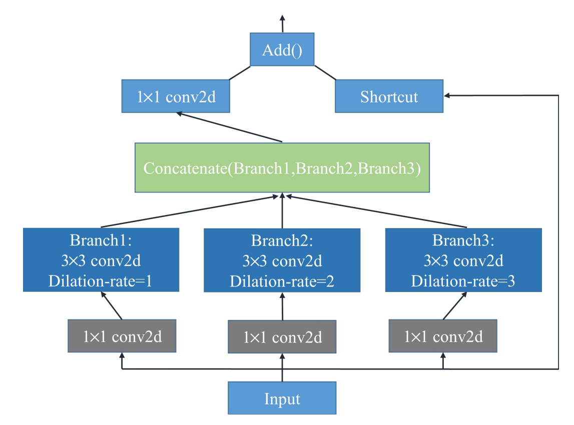
当前感受野大小如公式(3)所示。

$$F(i+1)=F(i)+(k'-1)\times L \qquad (3)$$

其中，$F(i+1)$ 为当前感受野大小；$F(i)$ 为上一层感受野大小；L 为不包括当前 $F(i+1)$ 层以外的前 i 个层步长之间的乘积。

Liu 和 Huang[20] 在 InceptionNet[21] 加入空洞

卷积，扩大感受野，增强网络的提取特征能力。借鉴其思路，建立并行的空洞卷积模块，称其为P_DConv（Parallel Dilated Convolution）。如图2所示。



注：Input：特征输入；1×1 conv2d：卷积核大小为(1，1)的卷积函数；Dilation-rate：空洞率；Concatenate：拼接函数

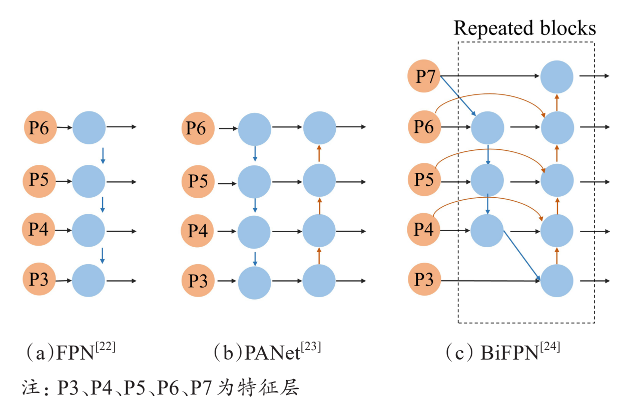
图2　并行空洞卷积结构

Fig. 2　Structure of parallel dilated convolution

输入的特征层通过1×1卷积降低通道数，经过3个不同空洞率的空洞卷积后将特征进行融合，组合Shortcut结构输出特征图。

2.4.2　特征金字塔

特征金字塔网络通过提取不同层的特征图，将高层与底层的信息融合形成多维度增强的特征。常见的特征金字塔结构如图3所示。



注：P3、P4、P5、P6、P7为特征层

图3　常见特征金字塔结构

Fig. 3　Common uses pyramid structures

参考文献［22—24］提出的结构，构建本研究的特征金字塔结构如图4所示。

F3至F6为特征提取网络输出的特征层，蓝

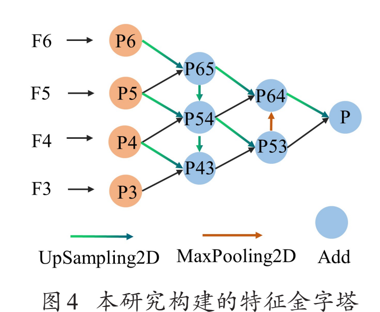


图4　本研究构建的特征金字塔

Fig. 4　The constructed feature pyramid in the research

色节点的执行顺序为P65、P54、P43、P53、P64、P。实现细节参照文献［24］，特征金字塔算法实现伪代码如下。

Algorithm　特征金字塔算法实现

```
Input：F3、F4、F5、F6
Output：P
def OurFPN(Input)：
  P3 = F3，P4 = F4，P5 = F5，P6 = F6，Features = P3，P4，P5，P6
  for j in range(len(Features))：
    C_Feature[j] = Conv2D(channels_num，kernel_size = 1) Feature[j]  #调整通道数
  P3_in = C_Feature[0]，P4_in = C_Feature[1]，P5_in = C_Feature[2]，P6_in = C_Feature[3]
  # P3_in、P4_in、P5_in、P6_in分别为通道数调整后的特征输入
  P6_UP = UpSampling2D()(P6_in)，P65 = Add([P6_UP，P5_in])，P65 = SeparableConv()(P65)
  # P6_in经过上采样与P5_in进行Add操作生成P65。P54、P43、P53生成操作相似
  P65_UP= UpSampling2D()(P65)，P5_UP = UpSampling2D()(P5_in)，P54=Add([P65_UP，P5_UP，P4_in])，P54= SeparableConv()(P54)
  # P65_UP和P5_UP经过上采样后与P4_in进行Add操作生成P54
  P54_UP= UpSampling2D()(P54)，P4_UP = UpSampling2D()(P4_in)，P43=Add([P54_UP，P4_UP，P3_in])，P43= SeparableConv()(P43)
  # P54_UP和P4_UP经过上采样后与P3_in进行Add操作生成P43
  P53 = Add([P54_UP，P43])，P53= SeparableConv()(P53)
  P53_MaP = MaxPooling2D()(P53)，P64 = Add([P53_MaP，P54，P65_UP])，P64= SeparableConv()(P64)
  # P53进行最大池化生成P53_MaP；P65经过上采样与P54、P53_MaP进行Add操作生成P64
  P64_UP = UpSampling2D()(P64)，P = Add([P64_UP，P53])，P= SeparableConv()(P)
  return P  #返回特征P，送入分类部分
```

将上述提出的并行空洞卷积和改进的特征金字塔结构与基于迁移学习的VGG16网络组合形成最终模型T-M-VGG（Transfer Learning-Multi-scale Feature Fusion-VGG），如图5所示。

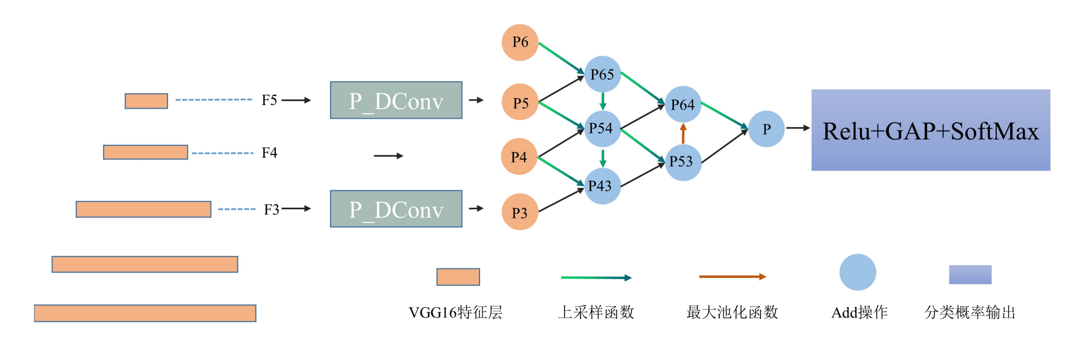


图5　T-M-VGG结构

Fig. 5　Structure of transfer learning-multiscale feature fusion-VGG

首先获取VGG16网络第三、第四和第五个卷积层输出的特征图F3、F4、F5，将F3、F5分别送入并行的空洞卷积模块P_DConv中。其次将P3、P4、P5、P6送入特征金字塔进行特征融合，P6由P5最大池化生成。最后把融合后的特征送入分类器输出结果。

2.5 试验参数设置

本研究的迁移学习实现方式为冻结预训练模型的全部卷积层，只训练自定义的全连接层，同时将全连接层更换为全局平均池化层。为更好地验证所提结构的有效性，采取对比试验形式，对比对象包括CNN结构[9-11,25]、VGG16[26]、MobileNetV3[27]（Large和Small两个版本）、InceptionV3[28]和人脸识别FaceNet结构（Inception-ResNetV2）以及采用预训练的基于迁移学习VGG16方法，其中基于迁移学习的VGG16记为Tr-L-VGG16，MobileNetV3（Large）记为Mb-Net-L，MobileNetV3（Small）记为Mb-Net-S。为防止模型训练发生过拟合，试验过程中所有方法采用早停机制，监控器为验证集的准确率val_acc；容忍因子min_delta=0.001；容忍步长patience=3。试验采取控制变量的方法，比对各个网络结构的性能，试验参数设置如表1所示。

2.6 试验评价指标

采用适合评价多分类任务的F1值(Macro_f1)和准确率(Accuracy)来评估模型的性能,各评价指标函数形式如下所示。

$$Macro_f1 = \frac{\sum_{i=1}^{N} f1_score_i}{N} \times 100\% \qquad (4)$$

其中，N为类别个数，个；$f1_score_i$为第i个类别的$f1_score$值。

$$Accuracy = \frac{Num_{TREUE}}{Num_{TOTAL}} \times 100\% \qquad (5)$$

其中，Num_{TREUE}为验证集中所有类别预测正确的个数，个；Num_{TOTAL}为验证集中所有样本的个数，个。

3 结果与分析

3.1 VGG系列算法结果与分析

早停机制使各个模型的训练及验证过程不同。图6和图7分别为不同试验方案在验证集生成的准确率和损失值结果。图例中T-M-VGG（train）表示经过训练集产生的结果；T-M-VGG（val）表示经过验证集产生的结果。

为使得模型评价角度多元化，将模型大小、可训练参数量作为辅助评价指标，结果见表2所示。

从图6和图7可以看出本研究所使用的各个

表 1　试验方案参数设置

Table 1　Parameters setting of experimental schemes

试验方案	图像训练形式	优化函数	学习率	批量	迁移学习-全连接层
文献[9]结构	256×256×3	SGD(momentum=0.9,decay=0.00001)	0.001	128	——
文献[10]结构	128×128×3	SGD(momentum=0.9,decay=0.00001)	0.001	128	——
文献[11]结构	256×256×1	SGD(momentum=0.9,decay=0.00001)	0.001	128	——
文献[25]结构	256×256×3	SGD(momentum=0.9,decay=0.00001)	0.001	128	——
VGG16	128×128×3	SGD(momentum=0.9,decay=0.00001)	0.001	128	——
Mb-Net-L	128×128×3	SGD(momentum=0.9,decay=0.00001)	0.001	128	——
Mb-Net-S	256×256×3	SGD(momentum=0.9,decay=0.00001)	0.001	128	——
InceptionV3	128×128×3	SGD(momentum=0.9,decay=0.00001)	0.001	128	——
FaceNet结构	128×128×3	SGD(momentum=0.9,decay=0.00001)	0.001	128	——
Tr-L-VGG16	256×256×3	SGD(momentum=0.9,decay=0.00001)	0.001	128	x=Activation('relu')(output) x=GlobalAveragePooling2D(x) x=Dense(194,activation='softmax')(x)
T-M-VGG	256×256×3	SGD(momentum=0.9,decay=0.00001)	0.001	128	x=Activation('relu')(output) x=GlobalAveragePooling2D(x) x=Dense(194,activation='softmax')(x)

注：将分辨率大小为256×256的数据集经过resize后得到128×128形式。图像训练形式中3代表RGB图像，1代表单通道图像。方案中VGG16、Mb-Net-L、Mb-Net-S、InceptionV3和FaceNet算法用ImageNet预训练模型参数初始化

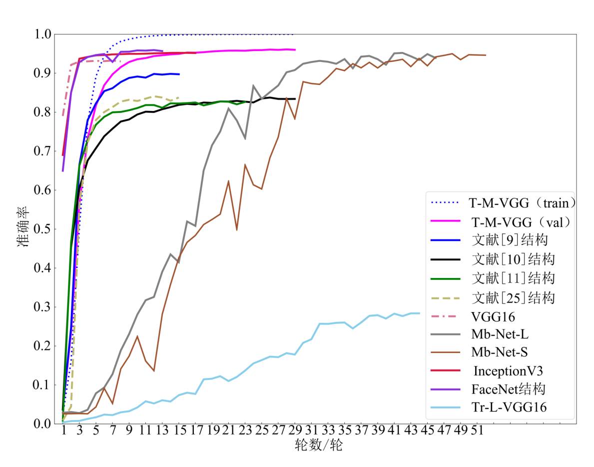


图 6　不同试验方案的准确率

Fig.6　Accuracies of different experiments

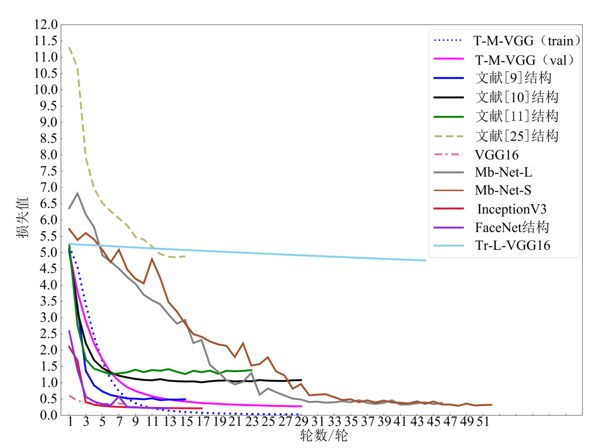


图 7　不同试验方案的损失值

Fig.7　Loss values of different experiments

方法随着迭代次数增加，准确率不断地升高，损失值不断地降低，达到一定次数后曲线趋于平缓。Tr-L-VGG16冻结卷积层后，自定义全局平均池化层参与训练的参数量较少，损失值下降缓慢，准确率曲线上升平缓，识别效果很差；VGG16采用预训练模型参数初始化后，准确率提升很快，迭代次数为7时达到收敛状态；文献［9］采用并行的迁移学习VGG16结构识别牦牛，在本试验环境下，迭代次数为14时接近收敛状态，损失值在0.5附近；从表2中可以了解，Tr-L-VGG16方法由于可训练参数量最小，所以准确率最低。VGG16经过训练、验证后模型大小为502.48 MB，所有方法中数值最高，不仅影响加载速度还增加存储开销。文献［9］模型大小为166.33 MB，准确率为88.03%，并行的迁移学习结构仍会增加模型大小。T-M-VGG模型大小为70.75MB，准确率为96.01%，对比VGG16、Tr-L-VGG16准确率分别提高近3个和68个百分点；对比文献［9］，牦牛识别方法准确率提高近

表2　不同试验方案的性能指标对照

Table 2　Comparison of performance indicators of different experiments

试验方案	F1值/%	模型大小/MB	准确率/%	可训练参数量/M
T-M-VGG	95.43	70.75	96.01	3.73
文献[9]结构	88.03	166.33	88.03	7.07
文献[10]结构	82.07	263.02	82.89	34.47
文献[11]结构	81.92	263.16	82.63	34.48
文献[25]结构	83.57	74.85	84.24	9.80
VGG16	92.85	502.48	93.02	65.85
Mb-Net-L	93.29	34.74	93.91	4.46
Mb-Net-S	94.60	13.65	94.62	1.72
InceptionV3	95.01	170.13	95.16	22.17
FaceNet结构	95.60	418.71	95.68	54.57
Tr-L-VGG16	20.64	56.96	28.38	0.10

8个百分点，模型大小减少约96 MB，这说明了并行空洞卷积和特征金字塔的有效性。

3.2 其他算法结果与分析

从图6可知，文献［10］、文献［11］和文献［25］中的网络结构在本研究超参数设定下走势相近，收敛状态时准确率保持在83%附近；FaceNet结构和InceptionV3网络收敛速度较快。从表2可以看出，T-M-VGG模型相比于文献［10］、文献［11］和文献［25］在准确率、可训练参数量两个方面保持着优势；对比InceptionV3和FaceNet结构，虽然准确率相差近1个百分点，但是模型占用空间分别减少了约100和348 MB。观察Mb-Net-L（MobileNetV3-Large）和Mb-Net-S（MobileNetV3-Small）算法，虽然模型较小，但是准确率低于T-M-VGG近2个百分点，收敛速度较慢。

综合上述分析，迁移学习方法能够显著地降低参数量，提升牦牛识别效果，特征金字塔结构能够提取丰富的牦牛面部特征进而完成识别过程。

3.3 识别效果可视化

为更好地检验所提模型的鲁棒性，从数据集中随机选取12张图像，对其部分区域进行遮挡操作生成伪图，将伪图送入T-M-VGG模型预测得到效果图，可视化结果如图8所示。

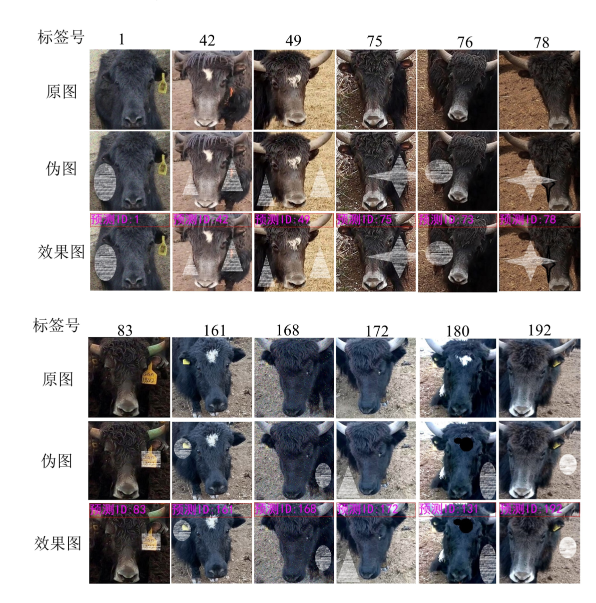


图8　T-M-VGG遮挡试验识别效果

Fig. 8　Recognition results of T-M-VGG for occlusion test

从标签号为1、42、49、76、83、161、168、172和192效果图可知，在遮挡牦牛非面部区域（身体，耳标，背景）的情况下，只有标签号76被预测为73，说明模型经过训练学习到的主要是面部特征，而非环境特征；从标签号为75和78效果图了解到，即使遮挡少部分脸部区域（非重要特征区域），模型仍然能够预测准确，但观察标签号为180的效果图，由于改变其面部明显特征，导致预测结果错误，这属于模型的正常表现。在12个类别中共有10个类别被T-M-VGG预测正确，准确率为83.33%。

4　结论与展望

本研究先后建立了并行的空洞卷积模块和改进的特征金字塔结构，结合迁移学习思想进行整合从而实现牦牛脸部识别算法，主要结论如下。

（1）提出的T-M-VGG模型在194头牦牛数据集中识别准确率为96.01%，模型大小为70.75 MB。

（2）在本研究构建数据集上通过和其他方法对比验证，说明了多尺度融合和迁移学习结合在

牦牛面部识别过程中的优越性。

（3）从准确率和模型大小等多个角度分析，采取输入分辨率大小为256×256的模型结构T-M-VGG能够在降低存储要求的同时提升准确率，基本满足实际的识别需求。

模型还存在不足之处，下一步工作将致力于以下几个问题的研究：（1）继续增加牦牛类别数量，扩大分类样本，对面部特征极为相似的牦牛识别进一步优化，提升模型的性能；（2）探究牦牛生长周期中体征变化对识别的影响；（3）结合目标检测算法完成实时识别。

参考文献：

[1] 秦兴, 宋各方. 基于双线性卷积神经网络的猪脸识别算法[J]. 杭州电子科技大学学报(自然科学版), 2019, 39(2): 12-17.
QIN X, SONG G. Pig face recognition algorithm based on bilinear convolution neural network[J]. Journal of Hangzhou Dianzi University (Natural Sciences), 2019, 39(2): 12-17.

[2] 何屿彤, 李斌, 张锋, 等. 基于改进YOLOv3的猪脸识别[J]. 中国农业大学学报, 2021, 26(3): 53-62.
HE Y, LI B, ZHANG F, et al. Pig face recognition based on improved YOLOv3[J]. Journal of China Agricultural University, 2021, 26(3): 53-62.

[3] 刘忠超, 何东健. 基于卷积神经网络的奶牛发情行为识别方法[J]. 农业机械学报, 2019, 50(7): 186-193.
LIU Z, HE D. Recognition method of cow estrus behavior based on convolutional neural network[J]. Transactions of the CSAM, 2019, 50(7): 186-193.

[4] 杨秋妹, 肖德琴, 张根兴. 猪只饮水行为机器视觉自动识别[J]. 农业机械学报, 2018, 49(6): 232-238.
YANG Q, XIAO D, ZHANG G. Automatic pig drinking behavior recognition with machine vision[J]. Transactions of the CSAM, 2018, 49(6): 232-238.

[5] 张宏鸣, 周利香, 李永恒, 等. 基于改进MobileFaceNet的羊脸识别方法研究[J/OL]. 农业机械学报: 1-10 [2022-05-13].
ZHANG H, ZHOU L, LI Y, et al. Sheep face recognition method based on improved mobilefacenet[J/OL]. Transactions of the CSAM, 1-10 [2022-05-13]. http://kns. cnki. net/kcms/detail/11.1964. S. 20220317.1251.014.html.

[6] 魏征. 基于全局和局部特征相结合的不完美牛眼虹膜识别技术研究[D]. 南京: 东南大学, 2017.
WEI Z. Research on iris recognition technology of imperfect bull's eye based on the combination of global and local features[D]. Nanjing: Southeast University, 2017.

[7] 赵凯旋, 何东健. 基于卷积神经网络的奶牛个体身份识别方法[J]. 农业工程学报, 2015, 31(5): 181-187.
ZHAO K, HE D. Recognition of individual dairy cattle based on convolutional neural networks[J]. Transactions of the CSAE, 2015, 31(5): 181-187.

[8] 何东健, 刘建敏, 熊虹婷, 等. 基于改进YOLOv3模型的挤奶奶牛个体识别方法[J]. 农业机械学报, 2020, 51(4): 250-260.
HE D, LIU J, XIONG H, et al. Individual identification of dairy cows based on improved YOLOv3[J]. Transactions of the CSAM, 2020, 51(4): 250-260.

[9] 陈争涛, 黄灿, 杨波, 等. 基于迁移学习的并行卷积神经网络牦牛脸识别算法[J]. 计算机应用, 2021, 41(5): 1332-1336.
CHEN Z, HUANG C, YANG B, et al. Parallel convolutional neural network yak face recognition algorithm based on transfer learning[J]. Journal of Computer Applications, 2020, 41(5): 1332-1336.

[10] HANSEN M F, SMITH M L, SMITH L N, et al. Towards on-farm pig face recognition using convolutional neural networks[J]. Computers in Industry, 2018, 98: 145-152.

[11] MARSOT M, MEI J, SHAN X, et al. An adaptive pig face recognition approach using convolutional neural networks[J]. Computers and Electronics in Agriculture, 2020, 173: ID 105386.

[12] KUMAR S, SINGH S K, SINGH R, et al. Deep learning framework for recognition of cattle using muzzle point image pattern[J]. Measurement, 2018, 116: 1-17.

[13] JUNG D H, KIM N Y, MOON S H, et al. Deep learning-based cattle vocal classification model and real-time livestock monitoring system with noise filtering[J]. Animals, 2021, 11(2): ID 357.

[14] SALAMA A, HASSANIEN A E, FAHMY A. Sheep identification using a hybrid deep learning and Bayesian optimization approach[J]. IEEE Access, 2019, 7: 31681-31687.

[15] WADA N, SHINYA M, SHIRAISHI M. Pig face recognition using eigenspace method[J]. ITE Transactions on Media Technology & Applications, 2013, 1(4): 328-332.

[16] RASHID M, GU X, LEE Y J. Interspecies knowledge transfer for facial keypoint detection[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2017: 1600-1609.

[17] WANG Z, BOVIK A C, SHEIKN H R, et al. Image quality assessment: From error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.

[18] 陈桂芬, 赵姗, 曹丽英, 等. 基于迁移学习与卷积神经网络的玉米植株病害识别[J]. 智慧农业, 2019, 1(2): 34-44.
CHEN G, ZHAO S, CAO L, et al. Corn plant disease recognition based on migration learning and convolutional neural network[J]. Smart Agriculture, 2019, 1 (2): 34-44.

[19] 李淼, 王敬贤, 李华龙, 等. 基于CNN和迁移学习的农作物病害识别方法研究[J]. 智慧农业, 2019, 1(3): 46-55.
LI M, WANG J, LI H, et al. Method for identifying crop disease based on cnn and transfer learning[J]. Smart Agriculture, 2019, 1(3):46-55.

[20] LIU S, HUANG D. Receptive field block net for accurate and fast object detection[C]// The European Conference on Computer Vision. Cham, Switzerland: Springer, 2018: 385-400.
[21] SZEGEDY C, WEI L, JIA Y, et al. Going deeper with convolutions[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2015: 1-9.
[22] LIN T Y, DOLLAR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2017: 2117-2125.
[23] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2018: 8759-8768.
[24] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2020: 10781-10790.
[25] 樊湘鹏, 周建平, 许燕, 等. 基于改进卷积神经网络的复杂背景下玉米病害识别[J]. 农业机械学报, 2021, 52(3): 210-217.
FAN X, ZHOU J, XU Y, et al. Corn disease recognition under complicated background based on improved convolutional neural network[J]. Transactions of the CSAM, 2021, 52(3): 210-217.
[26] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition[J/OL]. arXiv:1409.1556.2014.
[27] HOWARD A, SANDLER M, CHEN B, et al. Searching for Mobilenetv3[C]// The IEEE International Conference on Computer Vision. Piscataway, USA: IEEE, 2019: 1314-1324.
[28] SZEGEDY C, VANHOUCKE V, IOFFE S, et al. Rethinking the inception architecture for computer vision[C]// The IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2016: 2818-2826.

Multiscale Feature Fusion Yak Face Recognition Algorithm Based on Transfer Learning

CHEN Zhanqi[1], ZHANG Yu'an[1*], WANG Wenzhi[1], LI Dan[1], HE Jie[1], SONG Rende[2]

（1. *Department of Computer Technology and Application*, *Qinghai University*, *Xining* 810016, *China*;
2. *Animal Disease Prevention and Control Center of Yushu Tibetan Autonomous Prefecture*, *Yushu* 815000, *China*）

Abstract: Identifying of yak is indispensable for individual documentation, behavior monitoring, precise feeding, disease prevention and control, food traceability, and individualized breeding. Aiming at the application requirements of animal individual identification technology in intelligent informatization animal breeding platforms, a yak face recognition algorithm based on transfer learning and multiscale feature fusion, i.e., transfer learning-multiscale feature fusion-VGG(T-M-VGG) was proposed. The sample data set of yak facial images was produced by a camera named GoPro HERO8 BLACK. Then, a part of dataset was increased by the data enhancement ways that involved rotating, adjusting the brightness and adding noise to improve the robustness and accuracy of model. T-M-VGG, a kind of convolutional neural network based on pre-trained visual geometry group network and transfer learning was input with normalized dataset samples. The feature map of Block3, Block4 and Block5 were considered as F3, F4 and F5, respectively. What's more, F3 and F5 were taken by the structure that composed of three parallel dilated convolutions, the dilation rate were one, two and three, respectively, to dilate the receptive filed which was the map size of feature map. Further, the multiscale feature maps were fused by the improved feature pyramid which was in the shape of stacked hourglass structure. Finally, the fully connected layer was replaced by the global average pooling to classify and reduce a large number of parameters. To verify the effectiveness of the proposed model, a comparative experiment was conducted. The experimental results showed that recognition accuracy rate in 38,800 data sets of 194 yaks reached 96.01%, but the storage size was 70.75 MB. Twelve images representing different yak categories from dataset were chosen randomly for occlusion test. The origin images were masked with different shape of occlusions. The accuracy of identifying yak individuals was 83.33% in the occlusion test, which showed that the model had mainly learned facial features. The proposed algorithm could provide a reference for research of yak face recognition and would be the foundation for the establishment of smart management platform.

Key words: yak; face recognition; transfer learning; feature pyramid structure; T-M-VGG